CST1204: Introduction to Databases

Jun Shan
Adjunct Professor
City Tech, CUNY

Week 2 Session 1

9/4/2019

Agenda

- Teacher/Student Introduction
- Course Introduction
- Introduction to database concepts (Part 1 of 2)

Teacher/Student Introduction

- Name/Something special about yourself?
- How much do you know about IT and Database?
- What prompts you taking this course?

Course Description

- Content
  - Concepts and characteristics of relational database systems.
  - The organization of data including normalization and integrity constraints
  - Structure Query Language (SQL)

Course Communication

- Blackboard
  - Calendar
  - Content
  - Task
- OpenLab: hhttps://openlab.citytech.cuny.edu/shancst1204fall2019/

Class Organization

- Section E324: Mon/Wed 8-9:40p, Namm 1104
- School calendar on CityTech website:
  - Attached syllabus for weekly plan (subject to change)
  - Makeup class on Thu Sept 5th.
- Lecture + lab + test + homework. 2 class hours, 2 lab hours, 3 credits
- Textbook: Pratt, Philip J. and Mary Z. Last. 2009. A Guide to SQL, 9th edition. Boston: Cengage Learning, ISBN-13: 978-1-111-52727-3, ISBN-10: 1-111-52727-X

Grading

- Grading
  - 10% - Participation
  - 40% - Four Exams
  - 20% - Final Exam
  - 30% - Homework Assignments
- Participation is highly recommended
- Weather and emergency policy on City Tech website

Homework

- Homework released at the end of class and due before next class starts
- Two types
  - Q&A: Written in text format and email to instructor
  - Hands-on: Copy SQL as text and result as screenshots, paste into a Word document, and email to instructor.
- Email subject should be: “Homework - <your name> - <YYYYMMDD>”,where YYYYMMDD is the class date

Tests, Tests, and Final Test

- Four Tests: 1hr 40 min session. Will announce in the class before.
- Mid-term: Will provide each student a mid-term score around 10/30.
    - It is a mid-term feedback, not a grading.
- Final exam: 12/18. 1hr 40min. Grade will be available before 12/27.

Academic Integrity Policy

Students and all others who work with information, ideas, texts, images, music, inventions, and other intellectual property owe their audience and sources accuracy and honesty in using, crediting, and citing sources. As a community of intellectual and professional workers, the College recognizes its responsibility for providing instruction in information literacy and academic integrity, offering models of good practice, and responding vigilantly and appropriately to infractions of academic integrity. Accordingly, academic dishonesty is prohibited in The City University of New York and at New York City College of Technology and is punishable by penalties, including failing grades, suspension, and expulsion. The complete text of the College policy on Academic Integrity may be found in the catalog.

Hands-on and Lab Environments

- CUNY Oracle Server:
  - Remote Desktop
  - SQL Plus and SQL Workshop
- Oracle LiveSQL
- Students should have their own USB storage devices to store their homework/results

Database Concepts (Part 1 of 2)

- Relational databases
- Entity, Attribute, and Relationship

Database

- Database: a structure that contains different categories of information and the relationships between these categories. (Ch1. Pg 1)
- Enterprise data needs: Google, Facebook, Banks, Supermarkets
    - Data is essential to run business. Enterprises have lots of datasets.
    - Usually large and complex, related to each other in a specific way
    - Think of CUNY: Departments, teachers, students, listed courses, course registrations, etc.

Database Career

- Database is at the center of most enterprise computer applications. Database design and operation in many cases determines the computer application design and operation.
- Database is the foundation of Data Science, Artificial Intelligence, and Machine Learning
    - January 2019 report from Indeed showed a 29% increase in demand for data scientists year over year and a 344% increase since 2013

Database Management System

- The challenge is how to present the datasets in computer.
  - Entry, storage and retrieval of large quantities of information
  - Manage how that information is organized.
- The database management system (DBMS) is the software that interacts with end users, applications, and the database itself to capture and analyze the data.

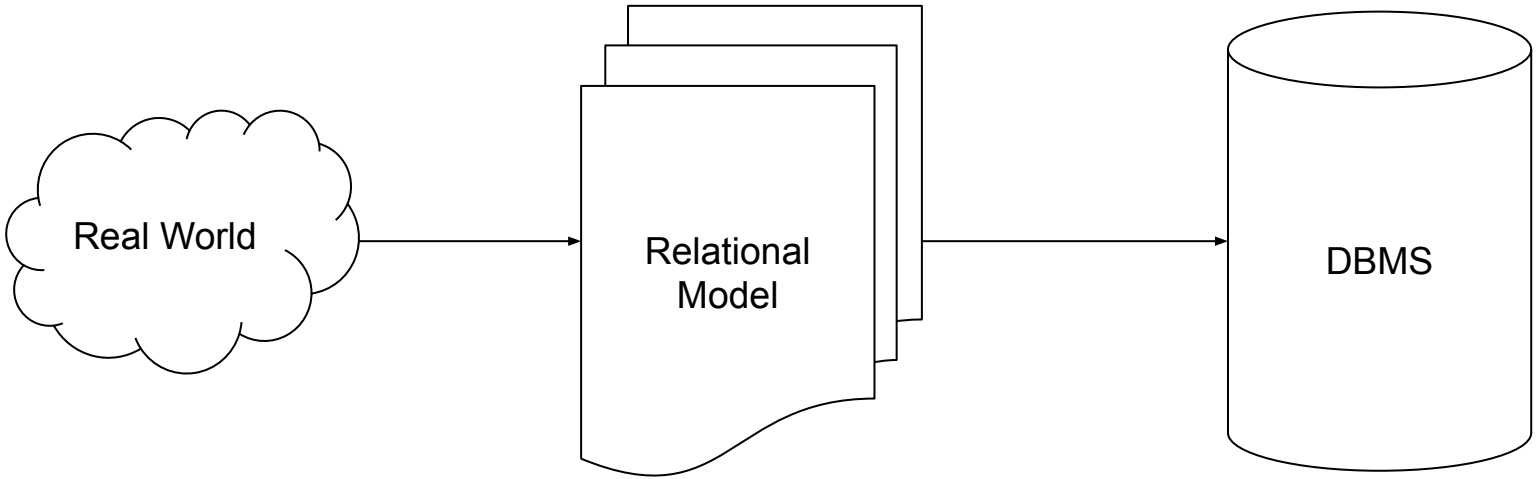
From File to Relation

- File system as DBMS (before 1970s):
  - Data duplication/loss.
  - No clear relationships between files. Hard to manage.
  - Security concerns
- Relational Model by Dr. Edgar F. Codd (1970):
  a. Even the most complex datasets can be represented as *Relations*.
  b. Relations are operated through Relational Algebra.

Relational Model

- Why is it important? Relational Model can be systematically, consistently, and logically represented in DBMS.
  - Has mathematical theory behind it (Relational Algebra)
  - Simple to develop and implement
  - Consistent in results across different platforms
  - Can be applied to almost all business scenarios

Relational Model



Relational Database

- Database that is built on Relational Model
- A collection of tables. Formally, these tables are called relations, and this is how this type of database gets its name. (Ch 2, Pg 22)
- Most commercial or open source DBMS on market today are RDBMS
    - Commercial: Oracle, SQL Server,
    - Open source: MySQL, PostgreSQL

Relational Database (RDB)

- Relational database examples: All 2-dimensional tables
  - Textbook example: Ch 2, Pg 22
  - Oracle Live SQL example
- How to create & organize these tables?
  - How to map real world questions into a relational database?

Relational Database Modelling

- How to create a relational database?
  - Logical Design: Through a systematic process of database modelling, we map real world questions into a logical model which contains entity, attribute, and relationships. The result is a Logical Model.
  - Physical Design: Map the entities and relationships in the Logical Model into database objects such as tables. The result is a Physical Model.

Entity, Attribute, and Relationship

- Entity: a noun; a person, place, thing, or event, such as a Sales Rep (Pg24).
- Attribute: a property of an entity, such as a Sales Rep Number.
- Relationship: the association between entities (Pg24). Usually a verb.
    - Different types of relationships
        - One-to-many: Sales Rep - Customer
        - Many-to-many:
        - One-to-one: State - State Capital
    - Relationships can have attributes (Descriptive Attribute)

Database Representation

- Relation: two dimensional tables in RDB (Pg26)
- Entity/Attribute: One entity per table, and attributes are columns (Pg22)
- Relationship: show the association between entities, aka putting entities in the same table.
    - Two direction linkage (Sales Rep - Customer)
    - Repeating group (Fig 2-2, Pg 25)
    - Must always be unique: Dups provide no new information

Entity Identification

- How to represent entity in a Relationship
  - How many columns do we need to represent an entity?
  - How to identify an entity instance?
- Shorthand representation (Pg27)
  - Qualified names

Hands-on Lab

- Initial survey
- Use shorthand representation to define entities, attributes, and relationships in TAL Distributor database as shown in Ch1.

Homework

- Watch the following video from Computer History Museum:

  https://www.youtube.com/watch?v=KG-mqHoXOXY